

平成22年5月14日現在

研究種目：基盤研究 (C)
研究期間：2007～2009
課題番号：19500201
研究課題名 (和文) エントロピーを導入した統計力学的ファジィクラスタリング手法の開発
及びその可視化
研究課題名 (英文) Entropy based statistical mechanical fuzzy clustering method and
its visualization

研究代表者
安田 真 (YASUDA MAKOTO)
岐阜工業高等専門学校・電気情報工学科・教授
研究者番号：80353275

研究成果の概要 (和文)：確定的アニーリング (DA) 法とファジィエントロピー及び Tsallis エントロピー最大化法を組み合わせたファジィ c 平均クラスタリング手法について研究を行った。これらのエントロピーとエントロピー最大化によって得られる帰属度関数は統計力学的な性質を持つため、その観点から分析・解釈を行った。

DA 法によるクラスタリングでは従来法を上回る分類性能を得た。また、超高速アニーリング法と帰属度関数の線形近似法の導入により、分類性能を落とさずにクラスタリング時間を短縮することに成功した。

クラスタリング手法を多次元のアンケートデータの分類に応用し、次元削除を行うことで、クラスタリング結果を効果的に可視化した。

研究成果の概要 (英文)：The fuzzy clustering methods which combine the deterministic annealing (DA) approach with fuzzy entropy or Tsallis entropy are examined. By maximizing fuzzy entropy or Tsallis entropy within the framework of the fuzzy c-means (FCM) method, membership functions similar to the statistical mechanical distribution functions are obtained. The characteristics of these entropies and membership functions are investigated from the statistical mechanical point of view.

Both fuzzy entropy and Tsallis entropy based FCM are formulated as DA clustering. Numerical experiments showed that these algorithms can divide data more properly and stably than the standard clustering method. The very fast annealing (VFA) method is applied to DA for cooling schedule of the temperature. Also, the membership functions are approximated by linear functions. The obtained results indicate that the proposed method can shorten computational times.

Multidimensional questionnaire data were clustered and, by omitting some dimensions, the resultant clusters and memberships were visualized.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	1,000,000	300,000	1,300,000
2008年度	500,000	150,000	650,000
2009年度	900,000	270,000	1,170,000
年度			
年度			
総計	2,400,000	720,000	3,120,000

研究分野：総合領域

科研費の分科・細目：情報学・感性情報学、ソフトコンピューティング

キーワード：ファジィ理論、情報統計力学

1. 研究開始当初の背景

統計力学は多数の要素からなるマクロな物理系の性質を調べる理論であるが、近年、情報処理の分野に統計力学的な方法を導入する試みが盛んとなり、情報統計力学と呼ばれる一分野を形成している。

一方、クラスタリングはデータに内在する構造を明らかにして分類する手法であり、データ解析の重要なツールである。実世界における多くのデータ分布ではクラスターの境界は必ずしも明確ではなく、クリスプなクラスタリングよりもファジィクラスタリングの方が適していると考えられる。ファジィクラスタリング手法としては、ファジィc-平均(FCM)法^[2]が標準的な手法となっている。我々はFCM法にファジィエントロピーを導入することで帰属度関数として量子統計力学のFermi-Dirac分布と同じ関数形が得られることを示し、ファジィクラスタリングのメカニズムやファジィエントロピーの統計力学的な解釈を行うとともに、シミュレーテッドアニーリング法や確定的アニーリング法などの緩和過程と組み合わせた新たなファジィクラスタリングアルゴリズムを提案した。

さらに、確定的アニーリング法とFermi-Dirac型の帰属度関数を組み合わせたアニーリングに基づくファジィクラスタリングについて調べ、相転移に基づくクラスター分割と分割終了条件を組み合わせたクラスター数の自動推定手法を提案し、ファジィクラスタリングの統計力学的な分析の有用性を示した。

2. 研究の目的

(1)ファジィエントロピーを導入したファジィクラスタリングと統計力学との対応関係

(粒子分布とデータ分布の対応や、分配関数と自由エネルギーの表現等)を明確にし、ファジィエントロピーとファジィクラスタリングのメカニズムの統計力学的な解釈を確立する。

(2) (1)で得た対応関係に基づく類推から、統計力学的な分布関数(及びエントロピー)と同型な新たなファジィクラスタリングの帰属度関数(及びエントロピー)を導出する。

(3)統計力学的な概念・物理量や分析手法のファジィクラスタリングへの導入や応用方法を検討する。

(4)Fermi-Dirac型の帰属度関数や(2)で得た帰属度関数とアニーリング法を組み合わせた新たなファジィクラスタリングアルゴリズムを開発し、数値実験による評価を行う。

(5) (4)で得たファジィクラスタリングアルゴリズムの実問題(主としてアンケートデータの分析)への応用評価を行う。

(6) (5)で得た応用結果の可視化処理について検討を行う。

3. 研究の方法

(1)Fermi-Dirac型の帰属度関数を用いたファジィクラスタリングアルゴリズムとして、確定的アニーリング(DA)法とシミュレーテッドアニーリング(SA)法の結合方法について検討する。ファジィクラスタリングの高速

化を実現するため、帰属度関数をより計算の容易な関数で近似して計算速度が向上が可能か数値実験により調べる。

(2) 統計力学的な帰属度関数を用いたファジィクラスタリングの特徴として、従来手法よりも複雑なクラスター形状が表現可能な点がある。そこで、(1)で開発したファジィクラスタリングアルゴリズムをアンケートデータ分析に応用し、複雑に分布するデータをファジィクラスタリングした場合の有効性を検証する。

(3) ファジィクラスタリングで得られる多次元の数値データをユーザに分かりやすく提供するため、数値データを人間にとって最も理解しやすい形態である画像データに変換することで、結果を分かりやすくユーザに提示する手法について調査・検討し、数値実験を行う。

4. 研究成果

(1) ファジィエントロピーを導入したファジィc平均法に基づくファジィクラスタリングと量子統計の関係について詳細に検討し、統計力学の世界とファジィクラスタリングの世界における分配関数、エネルギー、自由エネルギー、分布関数、エントロピーおよび粒子数一定の制約と規格化の制約などの対応関係を明らかにした。また、ファジィクラスタリングのメカニズムが粒子系の詳細釣合のアナロジーとして理解できることを示した。さらに、確定的アニーリング法と組み合わせたファジィクラスタリングでのファジィエントロピーの意味の解釈を示した。

(2) 遺伝的アルゴリズムを個体集団間の共進化メカニズムに基づいて拡張した競合共進化アルゴリズムでは、世代を重ねるうちに特定の遺伝子が優勢となって個体の多様性が失われるフォーカシング問題が生じる。そこで、ファジィクラスタリング手法により個体をその特徴に応じたクラスターに分類し、異なるクラスターから選択した個体を親個体とすることで競合共進化における多様性の維持に成功した。

(3) 確定的アニーリング法に基づくファジィクラスタリングにおいて、帰属度関数の線形近似により計算時間の短縮が可能であることを示した。

(4) 確定的アニーリング法に基づくファジィ

クラスタリングに適した温度の冷却法について検討し、従来法の線形法と指数法（超高速アニーリング法）を数値実験により比較した結果、超高速アニーリング法により高速かつ高精度なクラスタリングを行えることを示した。

(5) ファジィクラスタリングの目的関数をTsallisエントロピーで最大化することで統計力学的な帰属度関数が導出される。この帰属度関数と確定的アニーリング法を組み合わせることでファジィクラスタリングが実現される。数値実験の結果、本手法はファジィエントロピーを導入した従来手法と同様の性能を持つことを確認した。

(6) (5)の帰属度関数の性質について調べ、従来のファジィエントロピー最大化法の帰属度関数がガウス型の分布形状であるのに対し、新手法の帰属度関数はCauchy型であることを示した。Cauchy型のなだらかな分布形状により大域的なクラスタリングが可能になるという特徴を得た。

(7) Tsallisエントロピー最大化法と確定的アニーリング法の組み合わせにおいて、データ分布の広がりや帰属度関数の遷移領域幅の一致条件から初期温度が設定されることを示した。また、指数関数的に温度を下げる超高速アニーリング法の適用では、ファジィエントロピー最大化法よりも安定したクラスタリングが可能であることを確認した。

(8) 自己組織化マップはデータの可視化方法としてよく利用されるが、動的なデータの表現には適していないため、ノード交換により前後の時間の特徴を維持した特長マップの初期化法を提案し、数値実験により効果を確認した。

(9) ファジィクラスタリングの可視化手法として次元削除法を提案した。次元削除が帰属度に与える影響が最も小さい次元から順に削除することで高次元データの次元を削除し、帰属度の大小をデータポイントのサイズと色の濃淡で表すことで可視化を行った。アンケートデータのクラスタリング結果の可視化を行い、その効果を確認した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計6件)

- ① 安田真, 金井夏樹, 動的データにおけるノード交換を用いたSOM 特徴マップの初期化, 岐阜工業高等専門学校紀要, 査読有り, Vol.45, 2010 (掲載決定)
- ② 安田真, 高木雅也, ファジィエントロピーを利用したファジィc-平均法への確定的アニーリング法の適用, 査読有り, Vol.45, 2010 (掲載決定)
- ③ M.Yasuda, Fuzzy entropy based fuzzy c-means clustering with deterministic and simulated annealing methods, Institute of Electronics, Information and Computer Engineers Trans. Information and Systems, 査読有り, Vol.E92-D, 2009, pp.1232-1239
- ④ M.Yasuda, Deterministic and simulated annealing approach to fuzzy c-means clustering, IJICIC, 査読有り, Vol.5, 2009, pp.4981-4991
- ⑤ M.Yasuda, T.Furuhashi, S.Okuma, Statistical mechanical analysis of fuzzy clustering based on fuzzy entropy, Institute of Electronics, Information and Computer Engineers Trans. Information and Systems, 査読有り, Vol.E90-D, 2007, pp.883-888
- ⑥ 安田真, 辻嘉治, クラスタリングを用いた競合共進化手法の2人対戦ゲームへの適用, 岐阜工業高等専門学校紀要, 査読有り, Vol.42, 2007, pp.83-86

[学会発表] (計6件)

- ① 白木辰哉, 安田真, エントロピーを導入したファジィクラスタリング手法の開発, 平成21年度電気関係学会東海支部連合大会, 2009年9月10日, 愛知工業大学
- ② 外川達馬, 安田真, クラスタリングデータの可視化, 平成21年度電気関係学会東海支部連合大会, 2009年9月10日, 愛知工業大学
- ③ 金井夏樹, 安田真, 動的データにおけるノード交換を用いたSOM 特徴マップの初期化, 平成20年度電気関係学会東海支部連合大会, 2008年9月19日, 愛知県立大学
- ④ 高木雅也, 安田真, 確定的アニーリング法を適用したファジィエントロピー正則化ファジィc-平均法の冷却スケジュール, 平成20年度電気関係学会東海支部連合大会, 2008年9月1

9日, 愛知県立大学

- ⑤ M.Yasuda, Entropy maximization and deterministic annealing approach to fuzzy c-means clustering, International Conference on Soft Computing and 9th International Symposium on Advanced Intelligent Systems, September 19, 2008, Nagoya University
- ⑥ M.Yasuda, Deterministic and simulated annealing approach to fuzzy c-means clustering, International Symposium on Intelligent Informatics, December 12, 2008, Tokai University, Kumamoto Campus

[その他]

ホームページ等

<http://www.cc.gifu-nct.ac.jp/home/elec/yasuda/index.html>

6. 研究組織

(1) 研究代表者

安田 真 (YASUDA MAKOTO)

岐阜工業高等専門学校・電気情報工学科・教授

研究者番号: 80353275

(2) 研究分担者

無し

(3) 連携研究者

無し